Topics in Computational Inference

David M. Allen
University of Kentucky

October 11, 2018

6 Estimation

The subject of this Chapter is estimation of a linear combination of the elements of β in the linear model. The linear combination is represented by $c^t\beta$, where c is constant vector specified by the researcher. If there exists a linear combination of the response vector having expected value $c^t\beta$, then $c^t\beta$ is said to be *(linearly) estimable*. If X is not of full rank, not all of the elements of β are estimable.

The steps are

1. Determine if $c^t\beta$ is estimable.
2. If so, find its estimator and the standard error of the estimator.

Section 6.1 — General Concepts

Before proceeding, consider some examples.

Example 6.1. There are samples from two populations with respective means μ_1 and μ_2 and common variance σ^2. For inference on μ_1 choose $c^t = \begin{bmatrix} 1 & 0 \end{bmatrix}$, and for inference on $\mu_1 - \mu_2$ choose $c^t = \begin{bmatrix} 1 & -1 \end{bmatrix}$.

Example 6.2. Consider a regression model where the response variable is the change in systolic blood pressure. The first column of X is all ones for the intercept. The second column is age in days and the third column is measurements on birth weight. The expected change in systolic blood pressure associated with a unit increase in birth weight is β_1. For inference on β_1 choose $c^t = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$. To estimate systolic blood pressure of a three day old infant with birth weight 120 choose $c^t = \begin{bmatrix} 1 & 120 & 3 \end{bmatrix}$.

Method of Estimation

The method of estimation employed here is *minimum variance linear unbiased estimation*. This is closely related to *best unbiased estimators* discussed by Casella and Berger [1, Section 7.3.2].

Standard Error of the Estimate

For ease of expression, abbreviate $c^t\beta$ by θ. For inference, a measure of variability of an estimator is needed.

Definition 6.1. Given an estimator $\hat{\theta}$ for θ, the standard deviation of the distribution of $\hat{\theta}$ is called the *standard error* and is denoted $\mathrm{se}(\hat{\theta})$. The *estimated standard error* is denoted $\mathrm{ese}(\hat{\theta})$.

Section 6.2 — Estimators from a Transformed Model

Models having the basic variance structure, $\operatorname{Var} Y = \mathrm{I}\sigma^2$, have a simple and elegant analysis. These models may be written as

$$Y = X\beta + \epsilon.$$

Multiply from the left by an orthogonal matrix R^t to obtain

$$\begin{aligned} R^t Y &= R^t X\beta + R^t \epsilon \\ Y' &= X'\beta + \epsilon' \end{aligned} \tag{6.2.1}$$

The vector $\epsilon \sim N_n(0, \mathrm{I}\sigma^2)$ and $\epsilon' \sim N_n(0, R^t R\sigma^2 = \mathrm{I}\sigma^2)$ so the properties of the original model are preserved in the transformed model. The payoff comes when R is chosen so that X' has a simpler form than X.

Terminology

The maximum number of linearly independent columns of X is denoted $\text{rank}(X)$. If the columns of X are linearly independent then X is said to be *full rank*.

The Specific Transformation

A previous chapter gave a method for applying a sequence orthogonal transformations that transform a matrix X to echelon form. Let R^t represent the product of the transformation matrices. Then

$$R^t X = \begin{bmatrix} E \\ 0 \end{bmatrix}$$

where E is an echelon matrix. Partition R as $\left[R_X | R_\perp\right]$ such that $R_X^t X = E$ and $R_\perp^t X = 0$. Let $r = \mathrm{rank}(X)$ then R_X and $R_\perp$ have r and $n - r$ columns respectively.

Sub-Vectors of the Transformed Model

Sub-vectors of Y' in model (6.2.1) are extracted as

$$Y_X = R_X^t Y$$
$$Y_\perp = R_\perp^t Y$$

It follows from the distribution of linear combinations of normally distributed random variables that

$$Y_X \sim N_r(E\beta, \sigma^2 \mathrm{I})$$
$$Y_\perp \sim N_{n-r}(0, \sigma^2 \mathrm{I})$$

That $R_X^t R_\perp = 0$ implies Y_X and $Y_\perp$ are independent.

An Exercise

Exercise 6.1. Give the joint density function of Y_X and $Y_\perp$. Find minimal sufficient statistics for the parameters of model. Hint: See Casella and Berger [1, Theorem 6.2.13, Page 281].

Estimation of σ^2

Using the linear combination rule

$$\frac{1}{\sigma} Y_{\perp} \sim N_{n-r}(0, \mathrm{I})$$

so

$$\frac{Y_{\perp}^t Y_{\perp}}{\sigma^2} \sim \chi^2(n-r).$$

An unbiased estimator of σ^2 is

$$\widehat{\sigma^2} = Y_{\perp}^t Y_{\perp}/(n-r). \tag{6.2.2}$$

The Estimator and Its Standard Error

Consider the derivation of the minimum variance linear unbiased estimator for $\theta = c^t\beta$. The expression

$$L = a_X^t Y_X + a_\perp^t Y_\perp, \tag{6.2.3}$$

where a_X and $a_\perp$ are constant vectors, is a linear function of the elements of Y_X and $Y_\perp$. What does the unbiased property impose upon a_X and $a_\perp$? Inspection shows

$$\begin{aligned}\mathbb{E}(L) &= a_X^t \mathbb{E}(Y_X) + a_\perp^t \mathbb{E}(Y_\perp) \\ &= a_X^t E\beta.\end{aligned}$$

For L to be an unbiased estimator of $c^t\beta$, a_X must be such that $a_X^t E = c^t$ and $a_\perp$ can be anything at all. It could be that no such a_X exists, in which case $c^t\beta$ is not (linearly) estimable.

Assume for now that $c^t\beta$ is estimable. The rows of E are linearly independent, so a_X is unique. The variance of L is

$$\mathrm{Var}(L) = a_X^t a_X \sigma^2 + a_\perp^t a_\perp \sigma^2$$

Since a_X is fixed, $\mathrm{Var}(L)$ is minimized with $a_\perp = 0$. Since $a_\perp$ is no longer used, I rename a_X to a in what follows.

Summary

In summary, if there exist an a such that $a^t E = c^t$ then $a^t Y_X$ is the minimum variance linear unbiased estimator of $c^t \beta$. The variance of $a^t Y_X$ is $a^t a \, \sigma^2$ and its estimated standard error is $\sqrt{a^t a \, \widehat{\sigma^2}}$ where $\widehat{\sigma^2}$ is given in Equation (6.2.2).

Calculation of Y_X and $Y_\perp$

During the reduction of X to echelon form, information is saved that can be used to recreate the Householder transformations and apply them to Y. The matrices R_X and $R_\perp$ to need not be formed.

Computation of the Estimate and it Estimated Standard Error

If the solution of $a^t E = c^t$ for a exists,

$$\hat{\theta} = a^t Y_X$$
$$\mathrm{Var}(\hat{\theta}) = a^t a\, \sigma^2$$
$$\mathrm{ese}(\hat{\theta}) = \sqrt{a^t a\, \hat{\sigma}^2}.$$

If there is no solution, $c^t \beta$ is not estimable.

Independence of $\hat{\theta}$ and $\mathrm{ese}(\hat{\theta})$

The estimator $\hat{\theta}$ is a function of Y_X and $\mathrm{ese}(\hat{\theta})$ is a function of $Y_\perp$. Independence of Y_X and $Y_\perp$ implies independence of $\hat{\theta}$ and $\mathrm{ese}(\hat{\theta})$. See Casella and Berger [1, Page 161] for a discussion with respect to univariate random variables.

The cement data was first introduced as a textbook example by A. Hald in 1952 and has since been picked by several other authors including Tamhane and Dunlop [2]. It is shown in Table 6.1.

Exercise 6.2. For the cement problem, assume a model with an intercept and a linear function of the x_i. Find an estimate of the expected value of the third observation and the estimated standard error.

x_1	x_2	x_3	x_4	y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

Table 6.1: Hald's Cement Data

Section 6.3 — A Computational Method

This section gives a simple computational method, called the *tableau method*, for finding the estimate of $c^t\beta$ and its standard error. These quantities may be used to establish a confidence interval or to test a hypothesis. In the course of analyzing a model, we may consider several different c.

A Tableau Computational Method

The first step of implementation of the tableau method is forming the partitioned matrix

$$\left[\begin{array}{c} E \\ \hline c^t \end{array}\right].$$

For illustration, suppose that the numerical value or this matrix is

$$\begin{array}{rrrrr} 4 & -1 & 3 & 5 & -3 \\ 0 & 2 & 1 & -2 & 6 \\ 0 & 0 & 0 & 1 & 7 \\ \hline\hline 12 & -7 & 7 & 21 & -7 \end{array}.$$

The columns containing the first non-zero element in each row are highlighted.

Here is duplicate

$$
\begin{array}{rrrrr}
4 & -1 & 3 & 5 & -3 \\
0 & 2 & 1 & -2 & 6 \\
0 & 0 & 0 & 1 & 7 \\
\hline\hline
12 & -7 & 7 & 21 & -7
\end{array}.
$$

Looking at the first highlighted column, we see that subtraction of three times the first row from the last row would give a zero in the last row of that column. Save $(\alpha)_1 = 3$ and carry out the operation.

The result is

$$\begin{array}{rrrrr} 4 & -1 & 3 & 5 & -3 \\ 0 & 2 & 1 & -2 & 6 \\ 0 & 0 & 0 & 1 & 7 \\ \hline\hline 0 & -4 & -2 & 6 & 2 \end{array}.$$

Looking at the second highlighted column, we see that subtraction of -2 times the second row from the last row would give a zero in the last row of that column. Save $(\alpha)_2 = -2$ and carry out the operation.

The result is

$$\begin{array}{rrrrr} 4 & -1 & 3 & 5 & -3 \\ 0 & 2 & 1 & -2 & 6 \\ 0 & 0 & 0 & 1 & 7 \\ \hline\hline 0 & 0 & 0 & 2 & 14 \end{array}.$$

Looking at the third highlighted column, we see that subtraction of two times the third row from the last row would give a zero in the last row of that column. Save $(\alpha)_3 = 2$ and carry out the operation.

The result is

$$\begin{array}{rrrrr} 4 & -1 & 3 & 5 & -3 \\ 0 & 2 & 1 & -2 & 6 \\ 0 & 0 & 0 & 1 & 7 \\ \hline\hline 0 & 0 & 0 & 0 & 0 \end{array}.$$

We have $\alpha^t = \begin{bmatrix} 3 & -2 & 2 \end{bmatrix}$ such that

$$\alpha^t E = c^t.$$

and

$$\widehat{c^t\beta} = \alpha^t Y_1$$

Checking Estimability

If X is not of full rank then there are some c for which $c^t\beta$ will not be estimable. In the example, the highlighted elements in the last row zero were forced to 0, and the other elements in the last row also became zero. Had one of more of the non-highlighted elements not become zero, then $c^t\beta$ would not have been estimable.

Exercise 6.3. Let $n = 20$, $\hat{\sigma}^2 = 16$,

$$E = \begin{bmatrix} 4 & 6 & 8 & -1 \\ 0 & 3 & -2 & 4 \\ 0 & 0 & 2 & 3 \end{bmatrix} \text{ and } Y_X = \begin{bmatrix} 4.3 \\ 3.6 \\ 7.4 \end{bmatrix}$$

Find a 95% confidence interval for $c^t\beta$ when
$c^t = (4, 15, -2, 5)$.

References

[1] George Casella and Roger L. Berger. *Statistical Inference*. Second edition. Duxbury, 2002.

[2] Ajit C. Tamhane and Dorthy D. Dunlop. *Statistics and data analysis from Elementary to Intermediate*. Prentice Hall, 2000.